The Accuracy of the National Death Index When Personal Identifiers Other than Social Security Number Are Used

Brent C. Williams, MD, MPH, Lucy B. Demitrack, BA, and Brant E. Fries, PhD

ABSTRACT

This study analyzed the accuracy of the National Death Index when personal identifiers were used that included or excluded Social Security number. Computerized records of the Department of Veterans Affairs were used for comparison. Different combinations of identifiers other than Social Security number correctly identified from 83 to 92 percent of dead and 92 to 99 percent of living persons. These results should prove useful in ascertaining the mortality status of patient populations without information on Social Security numbers. (*Am J Public Health.* 1992;82:1145–1147)

Introduction

Accurate determination of mortality among subjects is critical for many types of research. To meet this need, the National Death Index has been established as a national computerized database of death records compiled by the National Center for Health Statistics.

Several studies have examined the accuracy of the National Death Index,[1–10] but all were completed during its first few years of operation, and few have focused on the comparative accuracy of different combinations of information that include or exclude Social Security number in correctly establishing death. Yet, for some populations information on Social Security number is not readily available. The usefulness of the National Death Index in determining the vital status of such groups depends on the accuracy of the matching process using information other than Social Security number.

We had the opportunity to compare the accuracy of the National Death Index using information that includes or excludes Social Security number during a study of health services use among users of facilities of the Department of Veterans Affairs.

Methods

The National Death Index

Details of the National Death Index application and matching process have been described elsewhere.[1,3,5] Briefly, the index contains information from all death certificates on file in state vital statistics offices since 1979, including decedents' names, Social Security numbers, dates of birth, and several demographic characteristics. A National Death Index user may specify which year(s) of death should be searched. Records are available for searches approximately 12 to 15 months after the end of a given calendar year. After approval and upon receipt of identifying information from the investigator, all cases from the National Death Index files that meet one or more of 12 selection criteria are appended and returned to the investigator for evaluation. The selection criteria have been described elsewhere[3] (and are available from National Death Index, Division of Vital Statistics, National Center for Health Statistics, 6525 Belcrest Road, Room 840, Hyattsville, MD 20782).

Data Sources and Variables

Data for our submission to the National Death Index were compiled from computerized hospital and nursing home records of all residents of Department of Veterans Affairs nursing homes between October 1986 and April 1987. Among other data requested, information for two groups of persons was submitted: (1) a random sample of persons recorded as dying in Department of Veterans Affairs hospitals or nursing homes through October 1, 1988 ("known dead," n = 500), and (2) all persons discharged alive or who received a clinical assessment in a Department of Veterans Affairs nursing home between September 15, 1988, and October 1, 1988 ("known alive," n = 3487). These data were matched with records from the National Death Index for the years 1986 through 1988, excluding index records with dates of death after October 15, 1988.

Analytic Methods

Department of Veterans Affairs records were assumed to reflect accurately patient mortality status. The analysis was

Brent C. Williams and Brant E. Fries are with the Ann Arbor Veterans Administration Medical Center Geriatrics Research Education and Clinical Center; Brent C. Williams is also with the Department of Internal Medicine, and Brant E. Fries with the Institute of Gerontology and School of Public Health, University of Michigan, Ann Arbor. Lucy B. Demitrack is with the Institute of Gerontology and School of Public Health, University of Michigan.

Requests for reprints should be sent to Brent C. Williams, MD, MPH, University of Michigan Medical Center, Department of Internal Medicine, Division of General Medicine, 3116 Taubman Center, Ann Arbor, MI 48109-0376.

This paper was submitted to the Journal October 27, 1991, and accepted with revisions April 21, 1992.

TABLE 1—Accuracy of Patient Identifiers in Ascertaining Death from the National Death Index

Identifiers	Identifiers Alone, %		Identifiers and Social Security Number, %	
	Sensitivity[a] (n = 500)	Specificity[b] (n = 3487)	Sensitivity[a] (n = 500)	Specificity[b] (n = 3487)
FI, LN, DB, MB, YB	83	99	83	99
FI, LN, DB, MB	85	97	85	99
FI, LN, MB, YB	88	93	88	99
FI, LN, YB	90	93	89	99
FI, LN, YB or LN, DB, MB, YB	92	92	92	99
Social Security number	...	...	97	99

Note. FI = first initial; LN = last name; DB = day of birth; MB = month of birth; YB = year (±1 year) of birth.
[a]Percentage of persons "known dead" with one or more National Death Index records that match on specified criteria.
[b]Percentage of persons "known alive" with no National Death Index records that match on specified criteria.

designed to determine the "sensitivity" and "specificity" of the National Death Index for ascertaining patient mortality. Sensitivity refers to the percentage of records of persons who were known dead that had at least one matching National Death Index record for a given set of identifiers. Specificity refers to the percentage of records of persons known alive that had no matching National Death Index records for a given set of identifiers. Sensitivity and specificity are collectively referred to as "accuracy."

A match on Social Security number was defined to occur if eight or nine digits were identical in the Department of Veterans Affairs and National Death Index records. A match on year of birth occurred if the two records differed by at most 1 year, an assumption that has been shown to improve the accuracy of matching.[5]

Results

Table 1 shows the accuracy of several combinations of identifiers in ascertaining the mortality status of members of our population, including or excluding Social Security number. Based on name and birth date alone, from 83% to 92% of known decedents and 92% to 99% of persons known to be alive were identified. The matching combination with the highest sensitivity (92%) was either (1) first initial, last name, and year of birth or (2) last name and complete (day, month, and year) date of birth. For investigators without Social Security number information, this combination would identify the greatest proportion of deaths with an acceptably high specificity (92%).

Among the known dead, the vast majority of "true" matches based on name and date of birth also matched on Social Security number. Thus, including information on Social Security number did not change the sensitivities substantially. However, of the observations from the known alive group that matched on name and birth-date identifiers, a substantial portion did not match on Social Security number. This was due to a relatively large number (n = 131) of (probably false) matches only on first initial, last name, and month and year of birth, and not on Social Security number. Including Social Security number as a matching criterion therefore significantly decreased the number of false-positive matches.

Social Security number alone had the best accuracy of any combination of identifiers examined, with a sensitivity of 97% and a specificity of 99%. The improved sensitivity of Social Security number alone over any other combination of name and birth-date identifiers was largely due to 13 cases among the known dead that (probably truly) matched on Social Security number and exact birth date, but not on last name. Matching on Social Security number alone thus yielded the highest number of true matches.

Adding information on race and/or marital status to the optimal combination of variables based on name and birth date did not significantly improve the overall results.

Discussion

These results demonstrate that among patient identifiers other than Social Security number, either first initial, last name, and year of birth or last name and full date of birth provides optimal ascertainment of death using the National Death Index. Matching on Social Security number alone, however, provided the best overall results. Therefore, investigators who wish to use the National Death Index to determine the vital status of a study population should include information on Social Security number whenever possible. If this is not possible, however, the National Death Index remains a highly accurate and useful resource and will correctly identify the mortality status of at least 92% of individuals using the criteria applied in this study.

Our estimates of the sensitivity and specificity of the National Death Index using identifiers other than Social Security number are probably conservative, for two reasons. First, inevitable misclassification of patients' mortality status in the Department of Veterans Affairs data would result in falsely low estimates. Second, the Department of Veterans Affairs data include only first initial rather than first name; complete information on first name would probably improve the specificity of the results without substantially decreasing the sensitivity.

This study had two important limitations: our data included few women and did not allow stratification of the results by race. Other studies[2,8] have suggested that the National Death Index may be less accurate for women and racial minorities than for other groups. The generalizability of these results to women and minorities should therefore be evaluated in future studies. □

References

1. Wentworth DN, Neaton JD, Rasmussen WL. An evaluation of the Social Security Administration Master Beneficiary Record File and the National Death Index in the ascertainment of vital status. *Am J Public Health.* 1983;73:1270–1274.
2. Curb JD, Ford CE, Pressel S, Palmer M, Babcock C, Hawkins CM. Ascertainment of vital status through the National Death Index and the Social Security Administration. *Am J Epidemiol.* 1985;121:754–766.
3. Edlavitch SA, Baxter J. Comparability of mortality follow-up before and after the National Death Index. *Am J Epidemiol.* 1988;127:1164–1178.
4. Boyle CA, Decouflé P. National sources of vital status information: extent of coverage and possible selectivity in reporting. *Am J Epidemiol.* 1990;131:160–168.
5. Patterson JE, Bilgrad R. The National Death Index experience: 1981–1985. In: Kilss B, Albey W, eds. *Proceedings of the Workshop on Exact Matching Methodologies, Arlington, VA, May 9–10, 1985.* Washington, DC: US Government Printing Office; 1985:245–254. US Department of the Treasury, Internal Revenue Service, publication no. 1299 (2-86).

6. Rogot E, Feinleib M, Ockay KA, Schwartz SH, Bilgrad R, Patterson JE. On the feasibility of linking census samples to the National Death Index for epidemiologic studies: a progress report. *Am J Public Health.* 1983;73:1265–1269.
7. Rogot E, Sorlie P, Johnson NJ. Probabilistic methods in matching census samples to the National Death Index. *J Chronic Dis.* 1986;39:719–734.
8. Acquavella JF, Donaleski D, Hanis NM. An analysis of mortality follow-up through the National Death Index for a cohort of refinery and petrochemical workers. *Am J Ind Med.* 1986;9:181–187.
9. Davis KB, Fisher L, Gillespie MJ, Pettinger M. A test of the National Death Index using the Coronary Artery Surgery Study (CASS). *Controlled Clin Trials.* 1985;6:179–191.
10. Stampfer MJ, Willett WC, Speizer FE, et al. Test of the National Death Index. *Am J Epidemiol.* 1984;119:837–839.

Race and Sex Differences in Mortality following Fracture of the Hip

Steven J. Jacobsen, MD, PhD, Jack Goldberg, PhD, Toni P. Miles, MD, PhD, Jacob A. Brody, MD, William Stiers, PhD, and Alfred A. Rimm, PhD

ABSTRACT

This paper describes the all-cause mortality experience, following a fracture of the hip, of 712 027 persons covered by the Medicare program from 1984 through 1987. White women experienced the lowest mortality rate (17.2 per 1000 person-months), followed by Black women (22.9 per 1000 person-months), Black men (33.5 per 1000 person-months), and White men (33.7 per 1000 person-months). The observed race-sex differences in survival were found at all ages and regardless of the number of comorbid conditions listed with the discharge diagnosis. While these data demonstrate marked race-sex differences in survival following hip fracture, the cause of these differences is not immediately apparent and demands further investigation. (*Am J Public Health.* 1992;82:1147–1150)

Introduction

Fracture of the hip represents a tremendous burden to the population aged 65 years and older.[1] White women experience the highest incidence of fracture, followed by White men, Black women, and Black men.[2–4] Following fracture, persons are at a much greater risk of death compared to the general population.[5–17] However, most studies of survival have been conducted in Scandinavia[6,11–13] and other predominately White populations,[14–16] and it is not known how well they generalize to the entire United States population or to minorities. The few studies that have included Blacks have been severely limited by a small sample size. The purpose of this study is to provide stable estimates of the survival following hip fracture of more than 700 000 Whites and Blacks aged 65 years and older in the United States.

Methods

Data for this study were obtained from the Health Care Financing Administration (HCFA) and have been described in detail elsewhere.[4,18,19] Briefly, HCFA maintains a computerized database of all hospital discharges covered by the Medicare program in the Medicare Provider Analysis and Review (MedPAR) file.[20] These data, keyed by Social Security Number, include the patient's age, race, and sex; up to five discharge diagnoses; zip code of residence; and dates of admission and discharge. To ascertain a date of death, records are cross-tabulated against death records from the Social Security Administration on a quarterly basis for two years following the date of admission; this follow-up information is incorporated in the summary file.

In the 1984 through 1987 MedPAR files, potential cases of hip fracture were identified as any individual with an ICD9-CM code[21] of 820.0 through 820.9 in any of the five fields provided for diagnoses (n = 810 949). Cases were excluded if age was less than 65 years (n = 19 899), if the fracture was the second or subsequent hip fracture during the four years (n = 54 055), if race was coded as other or unknown (n = 24 269), or if the fracture may have been secondary to neoplastic disease or the primary discharge diagnosis was for late effects of hip fracture (n = 4464). After exclusions, a total of 712 027 cases of hip fracture remained. (Some cases were excluded for more than one reason.)

The survival experience of persons following fracture of the hip was quantified

Steven J. Jacobsen is with the Department of Health Sciences Research, Section of Clinical Epidemiology, Mayo Clinic, Rochester, Minn. Jack Goldberg, Toni P. Miles, and Jacob A. Brody are with the Epidemiology Program, School of Public Health, University of Illinois at Chicago; Jack Goldberg is also with the VA Cooperative Studies Program Coordinating Center, Hines Hospital, Hines, Ill. William Stiers is with the Division of Rehabilitation Psychology and Neuropsychology, Department of Physical Medicine and Rehabilitation, University of Michigan Medical Center, Ann Arbor, Mich. Alfred A. Rimm is with the Division of Biostatistics/Clinical Epidemiology, Medical College of Wisconsin, Milwaukee, Wis.

Requests for reprints should be sent to Steven J. Jacobsen, MD, PhD, Department of Health Sciences Research, 200 First St SW, Mayo Clinic and Foundation, Rochester, MN 55905.

This paper was submitted to the Journal December 17, 1991, and accepted with revisions May 27, 1992.